图书情报工作
第65卷 第10期 2021年5月

智慧数据研究综述:概念辨析、价值取向、关键技术与应用框架*

■ 张云中 刘嘉琳
上海大学图书情报档案系 上海 200444

摘 要:[目的/意义]智慧数据是"智慧地球"背景下数据科学领域的新概念,目前其理论探讨与实践应用均在迅速发展。梳理学界对其认知脉络,凝聚共识、辨析差异,对厘清智慧数据理论体系和促进智慧数据应用开发意义重大。[方法/过程]在广泛深入阅读国内外领域相关文献的基础上,将智慧数据研究脉络梳理为概念内涵、价值取向、关键技术与应用框架四个分面,通过对比与分析,归纳出智慧数据的三种概念视角、五个特性特征、五类价值取向、三簇关键技术和五大应用领域。[结果/结论]研究发现,智慧数据的本质在于通过数据演进或结构设计实现数据规范精准及价值增值,其价值取向呈现出多元化复合特征,其技术体系旨在提供"可计算-可理解-可会话"的逐级演进支撑,其应用框架核心在于精准实现"数据"与"用户"的智慧交互,未来智慧数据理论体系还需在大"数据科学"观下围绕理论体系构建、数据权利治理、技术均衡发展、服务层级深化、理论实践交融等不断完善。

关键词:智慧数据 数据科学 价值 关键技术 应用 研究综述
分类号:G253
DOI:10.13266/j.issn.0252-3116.2021.10.014

1 引言

2008年IBM提出"智慧地球(Smart Planet)"项目,试图依托"互联网+物联网=智慧地球"的战略布局,实现人类生活方式的物联化、互联化与智能化,"Smart Data"正是从此中产生的新概念。"Smart Data"进入学术研究的视野后,学者们的认知出现分化:一种观点较为狭义,认为Smart Data是智能工具和技术的产物,意图通过先进的技术采集高流速与体量大的数字内容,从中挖掘超越性知识从而获得智慧,但往往又受限于数据技术水平未达到预期的智慧性[1],表现出"Smart,but not yet smart enough."的数据缺陷[2];另一种观点较为广义,认为Smart Data是数据形态和价值导向的产物,强调竭力避免现有数据形态的缺陷,破除数据量级的约束,力图结构规范、语义清晰、可视化与可对话的形态呈现数据价值。两类观点侧重不同,前者倾向于过程和技术,后者关注于结果和价值。随着大数据、云计算、人工智能等技术的发展与突破,两种观点逐渐交融,形成了借助相关技术分析和掌握数据、提炼体系化知识、升华出智慧并协助决策的数据的共识,并探索了丰富的智慧数据实践领域。当然,不容回避的是,学界关于智慧数据研究尚处在理论与实践交织的多元视角的探讨中,还没有形成较为清晰的认知脉络和理论体系:其是什么?特征如何?有何价值?通过何种技术可以实现?应用场景与前景如何?

截至2020年12月底,笔者以"智慧数据""Smart data"为主题词,在中国知网、Springer、Elsevier、Web of Science数据库中共检索出52篇中文文献与153篇外文文献,经过深入阅读,筛选整理出90余篇相关度较高文献。本研究立足智慧数据的发展脉络,旨在通过文献梳理和理论思辨,确定智慧数据的特性特征,概览智慧数据价值取向,归纳智慧数据关键技术,探究智慧数据应用框架,总结智慧数据发展模式与应用机理,以期全面综合地阐释智慧数据研究的最新观点与动态。

* 本文系教育部人文社会科学研究项目"公共数字文化资源展示的语义描述体系研究:'用户体验+数字人文'的视角"(项目编号:20YJA870022)研究成果之一。
作者简介:张云中(ORCID:0000-0002-7323-2561),副教授,博士,硕士生导师,E-mail:zhang-yun-zhong@126.com;刘嘉琳(ORCID:0000-0002-0158-9350),硕士研究生。
收稿日期:2020-12-25 修回日期:2021-02-08 本文起止页码:141-150 本文责任编辑:杜杏叶

2 智慧数据概念辨析

2.1 智慧数据的概念

智慧数据发展迄今,还未形成统一的概念及概念体系。目前学界对智慧数据概念的界定可粗分为三个视角:价值视角、结构视角、过程视角。笔者对不同视角下智慧数据定义做了相关整理,见表1。

笔者认为,价值视角下的智慧数据只是一张蓝图,强调了数据价值的递阶层级,但并未给出具体的方案;结构视角下的智慧数据侧重一步到位,强调对数据框架和模型的顶层精细化设计;过程视角下的智慧数据侧重逐步演进,强调自下而上的逐级提炼以创造新价值。三者之间各有侧重,统一于数据的规范精准及价值增值,尤以过程视角下"数字化——数据化——智慧化"三级演进的观点于现阶段被普遍接纳,数字化形成可计算的结构,便于利用与存储;数据化突出可理解的内涵,以语义化及可视化为进阶,旨在拓展智慧应用;智慧化呈现可会话的形态,表达具象价值,关注动态智慧的呈现形态[13]。

表1 不同视角下智慧数据的定义

定义视角	核心观点	代表学者及表述
价值视角	智慧数据是对事实抽丝剥茧后的高阶价值呈现	①从事实出发,将知识映射为Know-what、Know-how、Know-why,并结合丰富的真理呈现智慧的价值,呈现出"事实与观念-知识-智慧"结构[3];②数据有从低往高的价值提取与表现形式,呈现出"数据-信息-知识-智慧(Wisdom)"知识金字塔结构[4];③智慧数据是有价值的数据的子集[5]
结构视角	智慧数据经由规范设计而形成,形成智慧数据框架	①智慧数据较于一般数据更为"干净"[6];②智慧数据领域即是建立在给定数据集中构建能产生"智慧(Smartness)"的数据模型[7];③智慧数据形成阶段应该嵌入可解决4W1H(Who、Where、When、Why、How)疑难的智慧数据框架,从而权威描述智慧数据适用领域,框架越精化,描述适用领域内复杂对象的不确定性就越少[8]
过程视角	智慧数据经由演化路径形成	①智慧数据可由大数据转换[9];②"智慧数据"用于表示将原始数据转换为可以处理的数据,并以获得有价值的信息为目标[10];③智慧数据的建设需要经历"数字化——数据化——智慧化"阶段演进,从可计算、可理解走向可会话[11];④根据数据驱动的理念,智慧数据是由多源数据历经"数据化-语义化-可视化-智慧化"演化形成[12]

2.2 智慧数据的特点

目前学界对智慧数据的特性探讨尚没有体系化,仅有学者明确提出商业领域智慧数据必须具备的三个关键特性:准确性、可操作性和敏捷性[14]。在此基础上,笔者梳理了不同领域、不同维度智慧数据定义,从中归纳出智慧数据的五个特性:融合性、精确性、实用性、增值性、与决策性。

(1)融合性。智慧数据融合性是对数据来源、数据结构与数据适配性的融合。来源上需融合天地数据、社交数据、物流数据、办公数据、个人终端数据五大领域[15];结构上将杂乱、不可见的、未关联的、使用率较低的数据转化为规范组织、管理、可发现的、相互关联的、可重用的形态[5];适配性上需要适配特定领域、特定主题、特定行业的数据特性[16]。

(2)精确性。智慧数据精确性包含对数据内容描述的精准性要求和对数据自我保护形式的确定性要求。智慧数据的理想状态是拥有优质可靠的、精确真实的数据[17],有足够的精度驱动价值;同时要有足够的隐私保护与数据权力划分,诸如从纯数码格式转化为非明文数字字符串的模拟格式[18],从而规避可能的法律和技术问题,以合法合理地访问和使用数据。

(3)实用性。智慧数据的实用性是可操作性与敏捷性的结合。可操作性要求数据应推进可伸缩的操作,以实现跨平台应用等决策目标的最大化[14]。敏捷性要求数据必须是实时可用的,可随时准备好、灵活地适应不断变化的需求环境[14]。符合可操作性和敏捷性的智慧数据较之其他数据更为实用,具有"召之即来,来之能用"的优越性。

(4)增值性。智慧数据的增值性是对数据价值密度的提升,是高效率和高效用的要求。高效率要求数据收集、存储和分析等操作能力增强,以最大化利用数据[19]。高效用要求数据有更优质的语义表达,以更精炼的语义解释协助决策目标达成[20]。数据增值本质上是面向语义化、情景化的智慧数据逐级分析过程,其阶段性输出的知识成果是搭建知识金字塔的基石[21]。

(5)决策性。智慧数据决策性是前瞻性、系统性、评估性的协同。前瞻性指超越数据表象与现有知识水平的预见,理解并识别潜在影响的迹象[22]。系统性指在复杂和不确定环境下,结合资源配置与系统信息采取最佳行动的决策[23]。评估性包括对决策条件分析和决策后价值判断,旨在对动态变化、时效极强的数据作出准确评估[24]。

2.3 相关概念厘清

智慧数据与数据智慧:两者争议焦点是人的智慧

还是数据的智慧?数据本身无法获取意识,让数据识别人类意识与智慧将是一个战略性的错误[25],真正的智慧是人机互动的结果,人类的专业知识和创新思维才是智慧产生的内在动力[26]。

智慧数据与大数据:两者最大的区别在于数据量级和数据形态,大数据经预处理,使其能够被分析、处理、解释,并以结构化的方式访问它[27],即可转换为智慧大数据。在未来,大数据必然智慧化发展,智慧数据必然大数据化发展。

智慧数据与领域数据、科学数据:智慧数据统领各领域数据,搭建跨领域数据的数据桥梁[28]。科学数据是特定范畴的领域数据之一[16],部分科研数据具备智慧数据的雏形,具备基本的标准规范与质量控制。

3 智慧数据价值取向

智慧数据以价值为核心,处理原始数据以获得有价值的信息,尤指与决策相关的见解[29]。智慧数据价值取向旨在辨识智慧数据基本价值立场及所表现出来的基本价值态度,本研究将其归纳为管理价值、经济价值、创新价值、文教价值和法律价值。

3.1 智慧数据的管理价值

智慧数据管理价值旨在让智慧数据服务从正确地做事跃升为做正确的事。一方面,M. S. Javan、M. K. Akbari 从数据资源配置出发,认为智慧数据为数据采集、转换等操作提供数据融合抽象框架[30],使得多个有价值的数据能被跨平台观测与操作,从而提供更全面的决策愿景;另一方面,J. Chen 等学者从数据资源识别出发,认为智慧数据整合了针对性评估、可靠性分析、自动构建最优策略等实时交互功能[31],提供完全感知世界、动态了解事件、人机充分互动的智慧服务决策支持[32]。

3.2 智慧数据的经济价值

智慧数据经济价值是指挖掘湮没在海量数据中的经济价值[33],识别并消除数据处理过程中非增值环节,从而创造利润。在产品生产环节,S. Belkadi 指出智慧数据唤新了精益生产理念[34],以精细化的重要信息优化生产决策,最终指向本轻利厚的创收;在售后服务环节,英国政府调研智慧数据对消费市场的助力,发现借助智慧数据技术可实现消费者利益相关指标的远程采集、分析及反馈,找准基于客户需求的商业竞争优势,从而促进产品市场良性发展[35]。

3.3 智慧数据的创新价值

智慧数据创新价值体现在其可最大程度激活数据价值和知识潜能,推动信息技术的革新,增强人们创新创造的洞察力和决策力,从“已知-未知”走向“未知-未知”[27]。R. Kitchin 认为高级资源模式能满足科技创新对数据有效性、完整性、动态变化的极高需求[36],曾蕾指出通过关联数据、上下文感知、知识发现、情景重构等方式增强数据的表达能力[27],刘炜整理数字人文技术体系,指出智慧化服务将加速潜在研究项目的设计和发现[37]。

3.4 智慧数据的文教价值

智慧数据的文教价值是其在智慧文化中人文价值与智慧教育中育人价值的结合。对于文化资源,曾蕾指出智慧数据方法早被认可并应用在数字人文领域的智慧数据资源转换中,其目的是满足“互联网+文化场景”的文化服务需求[27]。对于教育教学,罗琳提出“数据-人-知识”的知识流动框架,其中智慧数据是能够成为知识酵母的数据,经过有效的知识转换与利用,智慧数据可智能化响应人的知识需求,成为为人所用的有价值的知识结合体[38]。

3.5 智慧数据的法律价值

智慧数据的法律价值是以人为本,控制数据隐私边界的安全保障。A. Cavoukian 以用户为中心提出智慧数据隐私框架,该框架以最终原则、合法原则、限制原则、透明原则、安全保障性原则、问责原则规范数据访问与使用操作[8]。同时,D. Roman 指出智慧数据将区块链、人工智能等技术应用在敏感数据保护中,主动将隐私和安全纳入数据自我保障范围[39],既实现数据透明使用又保障数据可追溯,是合法合理使用智慧数据的关键。

4 智慧数据关键技术

4.1 从智慧数据会议研究议题勘测智慧数据关键技术

智慧数据包含哪些关键技术?迄今,理论界和业界均没有给出权威的阐释。笔者认为,智慧数据技术是数字化、数据化、智慧化等几簇关键技术的集合,其本质是不断发展中的技术体系。本研究也尝试依据 IEEE 组织从 2015 年至 2020 年先后举办的 6 届 Smart Data Conference 的会议议题梳理智慧数据关键核心技术,通过对智慧数据演进路径中数字化(可计算)、数据化(可理解)、智慧化(可对话)三个环节的技术对比,探寻不同阶段智慧数据关键支撑技术的异同,如表 2 所示:

表 2 演进路径下智慧数据关键技术及各阶段数据形式

演进阶段	关键技术簇	具体技术类别	代表性技术方案与工具
数字化(可计算)	数据管理技术	数据采集与传输	物联网及 RFID;拍摄、采样、捕捉等;扫描;网络爬虫
		数据转换与集成	ETL 集成工具(Kettle、Talend、Apatar 等);NoETL 工具(Athena 等)
		数据存储	分布式文件/数据库系统;NewSQL/NoSQL(如 neo4j、JSON-LD 等)
		数据查询和索引	SQL 查询、SPARQL 查询
	数据安全技术	隐私认证	身份验证、虹膜识别、认证协议;
		去中心化技术	区块链技术、智能合同
	数据基础设施	开源智慧大数据系统	Hadoop、Spark、Flink 和 Storm 等
		智慧数据计算模式	边缘计算、云计算、并行计算、流计算等
数据化(可理解)	知识发现技术	数据挖掘/机器学习/深度学习	自动分类聚类、背景/情感分析、文本分析、主题分析、内容挖掘等
	语义化技术	语义网与语义本体	Protégé 等建模技术、本体映射及合并等
		关联数据	URI、XML、OWL、RDF(s)
		知识图谱	知识抽取、知识融合、知识推理等
	可视化技术	数据画像技术	用户画像、机构画像等
		可视化分析	可视化文本、图谱可视化、GPS/GIS 可视化、统计图表、三维建模、时空态势展示、VR/AR
智慧化(可会话)	认知技术	认知计算与计算神经科学	认知系统、人脑神经元、脑成像研究、脑启发计算、意图识别
		人工智能	遗传算法、人工神经网络、专家系统
		智能决策	自动化决策、大数据预测
	会话技术	自然语言处理	文本生成;句法分析、语义分析、语音识别
		智能问答技术	智能问答系统、搜索结果智能交互、推荐系统

4.2 智慧数据演进路径中的关键技术簇解析

(1)数字化关键技术簇解析。数字化旨在实现数据向可计算格式的转化,完成现实世界向数字世界的映射,构成智慧数据的数据基础。数字化的难点是在数据安全的前提下,集成分散数据与融合多源异构数据[40],完成数据的集成管理,夯实智慧数据基础设施建设。该阶段的关键技术包括:①数据管理技术:数据采集与传输,涉及利用物联网、采样等技术直接采集传输一手数据,或采用扫描、网络爬虫技术爬取网络平台的二手资料;数据转换与集成,包括重视转换流程的规范性的 Kettle、Talend、Apatar 等技术工具,或重视转换操作的可访问性的 Athena 等技术工具;数据存储,新兴的代表性技术如 Neo4j 等 NoSQL/NewSQL 数据库,可将数据直接存储在图结构中,以自由扩展、性能优异的架构优化存储资源;数据查询与索引技术,通过 SPARQL 语句对结构化的 RDF 数据做快速机器查询,系统性索引知识和技术。②数据安全技术:利用区块链等去中心化技术增强可信的分布式计算能力、建立安全可信的数据存储系统,维护数据的完整性、可用性;利用身份验证技术,通过虹膜验证、生物验证等手段完成身份管理、控制访问,维护数据的机密性。③数据基础设施:一类是为智慧数据流程运行提供平台基础的 Hadoop、Spark 等开源的智慧大数据系统;一类是为提升智慧数据计算效能的边缘计算、云计算、并行计算、流计算等新计算模式。

数字化关键技术最新进展主要聚焦在数据管理技术方面:数据管理的开发方案从围绕“数据存储”转变为围绕“数据安全”,NoSQL、NewSQL 数据管理系统虽已经取代关系型数据库成为对非结构化数据高效、可扩展分析的流行管理系统,但其演化原因是物理数据存储结构从键值对向图的转换[41],而在安全性的考量上不足,学界正对分布式存储与分布式处理技术的安全威胁作深入研究[42-43]。

(2)数据化关键技术簇解析。数据化旨在实现数据向语义化格式转化,提升数字内容的可理解程度并揭示其丰富的内涵。数据化的难点是避免因数据异构、语义缺失、知识隐藏等带来的数据分析、理解及可视化的困难,该阶段的关键技术包括:①知识发现技术:通过数据挖掘、机器学习、深度学习等方式展开自动分析,快速洞察多粒度数据的隐藏关系。②语义化技术:语义网[44]及本体是当前语义化技术的核心,目前呈现专深化发展态势,其催生了关联数据、知识图谱等新兴语义技术,实现了知识表示从面向“string”到面向“thing”的转变,奠定了智慧数据“可理解”的基础。③可视化技术:既涉及专门型的数字画像技术,可通过整合静态与动态特征数据实现用户或机构特征可视化;又涉及通用型的可视化手段,利用 VR/AR 技术实现多维度感知和虚拟交互[45],或利用 GPS/GIS 可视化技术实现数据三维可视化。

数据化关键技术最新进展主要集中在人机互动的

可视化技术领域：语义化技术标注语料从特定领域扩展到社会网络科学领域，标注对象主要面向图像，以图形理解、图像标注、图像检索等为热点研究[46－47]。可视化技术面向增强语义推理与降低视觉混淆交叉研究，从图形学基础走向交互视觉设计，从数据展示转换为数据探索[48,49]。

（3）智慧化关键技术簇解析。智慧化旨在形成可会话的人机交互模式，其难点在于训练机器模拟人的思维与认知，完成友好的人机交互模式下的智慧决策，拓展智慧应用的适用场景。该阶段的关键技术包括：①认知技术：传统技术包括利用遗传算法、人工神经网络、专家系统等算法或工具表示复杂的数据分布、内涵与特征，达到人工智能与智能决策的期望；新兴技术主要指认知计算与计算神经科学[50]，通过认知系统、人脑神经元等研究，解释人类认知行为与思维，为优化人机交互提供“接口”，升级计算机获取“智慧”的能力。②会话技术：一方面利用自然语言处理，通过文本生成表达自然语言意图，并用句法分析、语义分析正确理解知识真实含义，以高质量的自然语言处理迈向认知智能[51]。另一方面利用智能问答技术，通过智能交互、推荐系统识别人类的社交需求和情感暗示，完成深度推理、多轮交互问答等复杂自动问答[52]。

智慧化关键技术最新进展集中在会话技术方面：人机交互领域以人机对话技术为核心，在海量数据积累与深度学习技术的支持下，人机对话从基于符号规则和模板的对话系统、基于统计机器学习的对话系统迈向基于数据驱动的深度学习对话系统[53]。将其按照对话任务分类，面向特定领域或任务的对话系统向引入大规模常识方向发展，面向开放领域的非任务型对话系统向个性化对话方向发展[54]。

5　智慧数据应用框架

5.1　智慧数据应用机理

智慧数据应用是以智慧数据、智慧技术、智慧产品、智慧功能为对象边界的体系化服务，目前已拓展到智慧商务、智慧居行、智慧医疗、智慧文教、智慧科创等多个维度，各领域的典型应用在“依托何种数据资源、采用何种关键技术、形成何种数据形态、适用何种应用场景”四个关键问题上，呈现出较为明显的领域差异，如图 1 所示：

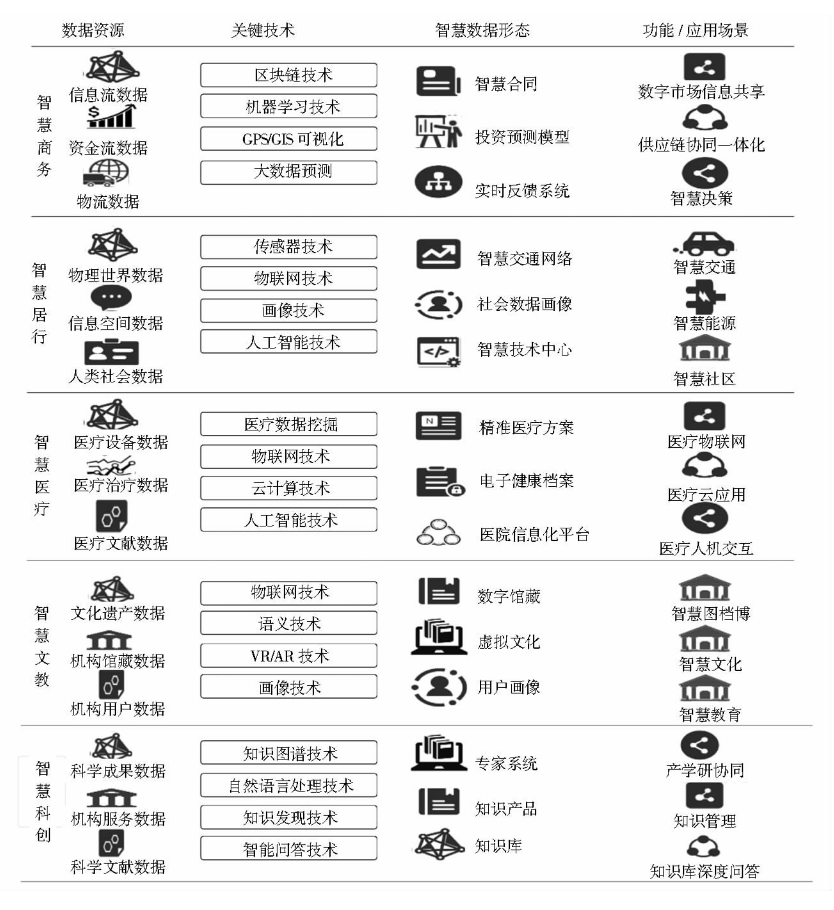


图 1　智慧数据应用框架

5.2 智慧商务

智慧商务是面向智慧数据驱动的商务模式,包含智慧金融、智慧物流、智慧商业等经济应用场景。智慧商务主要涉及信息流数据、资金流数据与物流数据,这些数据资源既散布在网络上,又集中在数据中心和云端[55]。其应用场景有:①应用区块链技术及其他安全认证技术,推广数字签名与智慧合同,增强移动支付的安全性、健壮性与私密性[56],简化金融资产交易流程。②应用机器学习技术分析金融信息,有效检测欺诈行为特征、预测金融、商业数据走势,支撑投资决策[57]。③应用 GPS /GIS 可视化技术,搭建智慧物流供应链构架,快速识别、定位、分拣、配送物流物料,使供应链过程可感、可视、可控[58]。③应用大数据预测技术,分析商务数据市场潜在的服务项目、隐含的营销思维,增加商务企业的盈利空间[59]。

5.3 智慧居行

智慧居行主要关注智慧城市建设中与衣食住行相关的多个方面,诸如智慧社区、智慧交通、智慧能源等;智慧居行需要对物理世界数据、信息空间数据、人类社会数据三元空间的数据进行语义细粒度融合。其应用场景有:①应用传感器技术,将各类基础设施融合成一个供给网络[60],促进社区治理中政府、企业、社会组织、民众等主体的互联互通[61]。②应用物联网技术,用三元空间数据构建智慧决策系统[62],并反向监测与规划智慧居行应用。③应用画像技术,标签化抽取社区特征并勾画社区数据画像[63],提供用户需求、环境监测、能源规划多种数据面板。④应用人工智能技术,深度融合自动化、机器化等技术以建设智慧技术中心[64],对智慧居行汇入的各种信息流做安全、隐私与风险管理[65]。

5.4 智慧医疗

智慧医疗以病人为中心,提供个性化专门医疗管理。智慧医疗要整理医疗设备、治疗方案以及医疗文献等庞杂的数据。其应用场景有:①应用物联网技术,构成医疗物联网[66],整合各医院医疗设备与医疗团队的数据,共建医疗资源共享。②应用医疗数据挖掘技术,通过深度学习医疗文献数据与临床有效诊疗数据的隐藏关系,有效筛查精准医疗方案。③应用云计算技术,研发医疗云应用,以云端存储的电子健康档案为核心,建立医院信息化平台;同时,突破医院与患者的时空限制实现云端治疗,节约就医诊疗双方的时间和经济开销;④应用人工智能技术,将人机交互的服务模式穿插进各种智慧医疗情境中,进一步完善电子健康档案,拓展医疗知识问答社区外延。

5.5 智慧文教

智慧文教旨在拓宽文化/教育机构馆藏的语义表达,增强文化/教育空间的全方位体验,为用户感知、学习文化/教育知识创设良好环境[67]。智慧文教需整合文化遗产与图档博等机构的馆藏数据、服务数据。其应用场景有:①应用物联网技术,实现智慧图档博等机构数字馆藏的提供,缩短数字空间中馆藏和用户之间的信息距离[68];②应用语义技术,结合馆藏特征、现有馆藏资源与用户需求,关联馆藏对象相关的解释性描述与多媒体内容外延,拓展馆藏能表达的数据内涵[69]。③应用 VR/AR 技术,融合虚拟与现实空间边界,提供文化遗产异质时空“对话”的互动体验,发展虚拟文化[70]。④应用画像技术,根据特定文化服务做主题融合,精准刻画具有时空属性的用户动态画像[71],不仅如此,在教育背景下,该技术可辅助测评师生互动、分析课程有效性、评定教学质量等,配置符合校园师生期待的教育资源,建设智慧教育环境[72]。

5.6 智慧科创

智慧科创是知识跨组织、跨领域流动的交叉学习过程,通过协同创新提升智慧数据创新价值。智慧科创需其广泛集成文献数据、科研成果数据、机构服务数据等内容资源。其应用场景有:①利用知识图谱技术,以图谱形式直观展示科学知识之间的关联与结构,发现其中的规律并揭示其发展演进的概貌[73]。②利用自然语言处理技术,通过机器翻译创造快速阅读的条件,通过文本生成自动撰写高质量数据报告,减负人工工作[74]。③应用知识发现技术,挖掘、聚类、分析用户行为、社会关系等数据内容,探寻深层次信息的启示,改善知识管理[75]。④应用智能问答技术,升级可用于情报分析的数据库、知识平台等知识产品,基于用户提问做信息检索与深度问答,满足用户个性化的检索与分析需求[76]。

6 结论与讨论

本研究对国内外智慧数据领域目前取得的相关研究成果进行了系统化梳理从概念辨析、价值取向、关键技术、应用框架四个基础维度,并深层次探讨发展模式、应用机理,共归纳如下六点关键结论:

(1)智慧数据的思想可溯源至数据递阶模型,概念内涵的核心在于通过数据演进或结构设计实现数据规范精准及价值增值。智慧数据具有融合性、精确性、实用性、增值性、与决策性,其“智慧”是人机互动的结

果。

(2)智慧数据以价值为核心,其价值立场和价值取向呈现出多元化复合特征,着重体现在管理价值、经济价值、创新价值、文教价值和法律价值等五个向度。智慧数据的价值取向是智慧数据应用的价值和效用前提。

(3)智慧数据关键技术体系是数字化、数据化、智慧化等几簇关键技术的集合,本质是为智慧数据"可计算－可理解－可会话"演进路径提供技术支撑,其中的技术类别及代表性技术方案都并非是固化的,而是不断动态更新的。

(4)目前智慧数据典型应用包含但不限于智慧医疗、智慧居行、智慧商务、智慧文教、智慧科创等领域,均属于智慧城市范畴下的细分领域。智慧数据的应用框架旨在解决如何精准实现"数据"与"用户"的智慧交互问题,涉及的核心在于:哪些采集数据源?数据如何演进?智慧数据呈现为何种形态?有何功能及应用场景?

(5)智慧数据在理论研究和实践探索的双向交融下,逐步形成了深入呈现数据资产价值的智慧数据发展模式:在数据结构上,从侧重"逐级演进"逐步发展为侧重"顶层设计";在数据处理上,从注重后台"数据分析"逐步发展为前端注重"会话交互";在技术支撑上,从数据科学"通用技术"逐步孵化出智慧数据"专门技术";在应用领域上,从"经典领域"正拓展为"全部行业";在应用能级上,从"智慧运营"逐步迈进为"智慧决策"。

(6)智慧数据的应用机理如图2所示,①问题/用户是价值创造的始发点,他们提出了包含服务需求、产品需求等不一而足的重要需求;②场景/功能是价值增值的中转站,其具象出智慧产品的雏形与核心功能,并确定功能实现所需数据的演化形态与模式;③资源是价值加工的目标对象,加工过程为通过智慧数据技术演化出优质、规范的智慧数据,用于匹配场景/功能的运作需求,进一步满足初始问题/用户需求。



图2　智慧数据应用机理

诚然,在梳理前人研究过程中,笔者也发现了当前研究尚存的局限性或不足,值得后续研究中继续深入探讨:

(1)大"数据科学"观下的智慧数据理论体系研究。智慧数据是数据科学领域的新生概念,其理论体系正在不断形成,建构尚不完善,且与大数据、科学数据、领域数据、开放数据等数据科学其他领域既有交集,更有区分。未来的研究一方面应突出智慧数据作为一种对数据全新认知的方式,其理论体系上的独有性,另一方面也应强调智慧数据与大数据等领域的高度关联性,探索智慧数据为核心的数据科学交叉融合研究。

(2)从"价值"探讨进一步深化拓展到"权利"探讨。当前研究视角多聚焦在价值探讨上,多在思考智慧数据有何用,而对智慧数据使用的合法性未给予充分重视。后续研究应进一步拓展智慧数据权利探讨,开展智慧数据权利治理相关研究,思辨智慧数据的权利的生成逻辑和构成主张,探寻各领域内数据安全、数据脱敏等权利保障的技术及制度手段。

(3)智慧数据技术体系应均衡"逐步演进"类与"规范设计"类。尽管目前学界形成了智慧数据"由演进而来"和"由设计而来"两种认知,但聚焦到技术体系上,前一观点更被广泛接纳,这也是目前智慧数据技术体系呈现出"数字化、数据化、智慧化"技术工具居多的根源所在。未来的智慧数据技术体系,需要不断吸纳数据规范设计视角下的智慧数据新技术与新工具,使得智慧数据的产生能够"一步到位",最终形成"逐步演进"与"规范设计"均衡发展的综合技术体系。

(4)智慧数据应用领域还应进一步提升。智慧数据是应用型的数据科学,目前其应用领域不断深化,已涉及商务、交通、医疗、文化、科创等多个领域,但现阶段其应用层级以面向智慧运营居多。未来智慧数据应用,应提供精准匹配、智慧交互、互利共赢的用户体验和服务效能,实现从面向智慧运营到面向智慧决策的能级提升,形成全方位、跨平台的高质量智慧应用。

参考文献:

[1] BALAKRISHNA S, SOLANKI V K, GUNJAN V K, et al. Performance analysis of linked stream big data processing mechanisms for unifying IoT smart data[C]// GUNJAN V, GARCIA D V, CARDONA M, et al. ICICCT 2019 - system reliability, quality control, safety, maintenance and management. Singapore: Springer, 2020: 680－688.

[2] SEIFERT A, HOFER M, ALLEMAND M. Mobile data collection: smart, but not (yet) smart enough[EB/OL]. [2020－12－24]. https://doi.org/10.3389/fnins.2018.00971.

[3] CLEVELAND H. Information as a resource[J]. Futurist, 1982, 16(6): 34－39.

[4] ACKOFF R L. From data to wisdom[J]. Journal of applied systems analysis, 1989, 16(1): 3-9.

[5] IAFRATE F. From big data to smart data[M]. London: ISTE Ltd, 2015: 9.

[6] SCHÖCH C. Big? smart? clean? messy? data in the humanities [EB/OL]. [2020-12-24]. http://journalofdigitalhumanities.org/2-3/big-smart-clean-messy-data-in-the-humanities/.

[7] FRANCE R, RUMPE B. Modeling big smart data[J]. Software and system modeling, 2014, 13(2): 457-458.

[8] CAVOUKIAN A, TOMKO G, BORRETT D, et al. The security architecture discipline as catalyst to effect secure smartdata design: employing abuse case elaboration for attack vector discovery and countermeasure determination[M]// HARVEY I, CAVOUKIAN A, TOMKO G, et al. Smartdata. New York: Springer, 2013: 173-185.

[9] MICHAEL F. Turning big data into smart data[EB/OL]. [2020-12-24]. https://sloanreview.mit.edu/article/turning-big-data-into-smart-data/#article-authors.

[10] LENK A, BONORDEN L, HELLMANNS A, et al. Towards a taxonomy of standards in smart data[C]//HO H, OOI B, ZAKI M J, et al. 2015 IEEE international conference on big data. New York: IEEE, 2015: 1749-1754.

[11] 王晓光. 数字人文与智慧数据[J]. 上海高校图书情报工作研究, 2018, 28(2): 24-25.

[12] 郑建明. 专题:智慧公共文化服务研究[J]. 图书馆论坛, 2020, 40(9): 13-13.

[13] HARVEY I. Perspectives on artificial intelligence: three ways to be smart[M]// HARVEY I, CAVOUKIAN A, TOMKO G, et al. Smartdata. New York: Springer, 2013: 27-38.

[14] GARCIA-GIL D, LUENGO J, GARCIA S, et al. Enabling smart data: noise fifiltering in big data classifiication[EB/OL]. [2020-12-24]. https://arxiv.org/abs/1704.01770.

[15] 顾立平,陈新兰,张潇月,等. 开放科研数据中的数据价值提升策略[J]. 图书馆论坛, 2020, 40(9): 115-124.

[16] 李阳,孙建军. 人文社科专题数据库建设规范化管理的若干问题[J]. 现代情报, 2019, 39(12): 4-10.

[17] LUENGO J, GARCÍA-GIL D, RAMÍREZ-GALLEGO S, et al. Final thoughts: from big data to smart data[M]//JULIÁN L, DIEGO G, SERGIO R, et al. Big data preprocessing. Cham: Springer, 2020: 183-186.

[18] CHOENNI S, BARGH M, ROEPAN C, et al. Privacy and security in smart data collection by citizens[J]. Smarter as the new urban agenda, 2016(11): 349-366.

[19] MARR B. Why only one of the 5 Vs of big data really matters[EB/OL]. [2020-12-24]. https://www.ib mbigdatahub.com/blog/why-only-one-5-vs-big-data-really-matters/.

[20] TRACTENBERG R E, SELLERS K F. "Small" Data[EB/OL]. [2020-12-24]. https://doi.org/10.1007/978-3-319-32001-4_445-1.

[21] KITCHIN R. The data revolution: big data, open data, data infrastructures and their consequences [M]. THOUSAND OAKS. California: Sage Publications, 2014.

[22] IAFRATE F. A journey from big data to smart data[C].//KROB D, LONJON A, BENGHOZI P J, et al. Proceedings of the second international conference on digital enterprise design and management DED&M 2014. Cham: Springer, 2014: 25-33.

[23] TRUNK A, BIRKEL H, HARTMANN E. On the current state of combining human and artificial intelligence for strategic organizational decision making[J/OL]. [2020-12-24]. https://rd.springer.com/article/10.1007%2Fs40685-020-00133-x.

[24] CUZZOCREA A, GRECO S. Flexible querying and analytics for smart cities and smart societies in the age of big data: overview of the FQAS 2019 international conference[C] //CUZZOCREA A, GRECO S, LARSEN H, et al. Flexible query answering systems. Cham: Springer, 2019: 25-28.

[25] TOMKO G, BORRETT D, KWAN H, et al. Smartdata: make the data "think" for itself[J]. identity in the information society, 2010, 3(2): 343-362.

[26] 王涛. "数字史学":现状、问题与展望[J]. 江海学刊, 2017(2): 172-176.

[27] 曾蕾,王晓光,范炜. 图档博领域的智慧数据及其在数字人文研究中的角色[J]. 中国图书馆学报, 2018, 44(1): 17-34.

[28] 杨林霞. 面向科研用户的嵌入式智慧数据服务模式研究[J]. 图书馆, 2019(10): 28-33.

[29] ELAGGOUNE Z, MAAMRI R, BOUSSEBOUGH I. A multi-agent framework for medical diagnosis driven smart data in a big data environment[C]// AUER M, TSIATSOS T. Proceedings of the 11th IMCL conference. Cham: Springer, 2018: 720-727.

[30] JAVAN M S, AKBARI M K. SmartData 4.0: a formal description framework for big data[J]. Journal of Supercomputing, 2019, 75(7): 3585-3620.

[31] CHEN J, DOSYN D, LYTVYN V, et al. Smart data integration by goal driven ontology learning[C]//ANGELOV P, MANOLOPOULOS Y, ILIADIS L, et al. Proceedings of the 2nd INNS conference on big data. Cham: Springer, 2016: 283-292.

[32] DANH L P, HOAN N M Q, HUNG N Q, et al. The graph of things: a step towards the live knowledge graph of connected things[J]. Journal of web semantics, 2016(37/38): 25-35.

[33] ZHANG C, CHEN D, TAO F, et al. Data driven smart customization[C]//BUTALA P, GOVEKAR E, VRABIC R. Proceeding of 52nd CIRP conference on manufacturing systems. Amsterdam: Elsevier, 2019: 564-569.

[34] SOUKAYNA B, ILIAS C, MOHAMED B. Lean in information technology: produce the human before the software[C]// Khoukhi F. Proceedings of the international conference on advanced information technology, services and systems (AIT2S-18). Cham: Springer, 2019: 273-283.

[35] GREG C. Smart data putting consumers in control of their data and

enabling innovation[EB/OL]. [2020 – 12 – 24]. https://www.gov. uk/government/consultations/smart-data-putting-consumers-in-control-of-their-data-and-enabling-innovation.
[36] KITCHIN R, LAURIAULT T P. Small data in the era of big data [J]. Geojournal, 2015, 80(4): 463 – 475.
[37] 刘炜,叶鹰. 数字人文的技术体系与理论结构探讨[J]. 中国图书馆学报, 2017, 43(5):32 – 41.
[38] 罗琳,顾新. 智慧数据驱动的产学研协同创新知识管网研究[J]. 软科学,2017,31(6):15 – 18.
[39] ROMAN D, VU K. Enabling data markets using smart contracts and multi-party computation [M]//ABRAMOWICZ W, PASCHKE A. Business information systems workshops. Cham: Springer, 2019: 258 – 263.
[40] 张杨燚,谢辉,毛进,等. 面向城市数据画像构建的多源数据需求与融合方法研究[J]. 情报理论与实践, 2020, 43(6): 88 – 96.
[41] CHAUDHRY N, YOUSAF M M. Architectural assessment of NoSQL and NewSQL systems [J]. Distributed parallel databases, 2020,38(4):881 – 926.
[42] CHAUDHRY N,YOUSAF M M, KHAN M T. Security assessment of data management systems for cyber physical system applications [J]. Journal of software: evolution and process, 2019, 32(3): ee2241.
[43] CASTONGUAY J, STEIN S S. Digital assets and blockchain: hackable, fraudulent, or just misunderstood[J]. Accounting perspectives, 2020, 19(4): 363 – 387.
[44] 许鑫,杨佳颖. 国外语义网研究现状与动向——基于 2002 – 2018 年 ISWC 会议[J]. 情报学报, 2020, 39(7): 761 – 776.
[45] LANDI D. The image of the hyper city[J]. International journal for the semiotics of law, 2019, 32(3): 533 – 548.
[46] LIETZ H. Drawing impossible boundaries: field delineation of social network science [J]. Scientometrics, 2020, 125(3): 2841 – 2876.
[47] LONG C, YANG X, XU C. Cross-domain personalized image captioning[J]. Multimedia tools and application, 2020, 79(45 – 46): 33333 – 33348.
[48] PICCIALLI F, CASOLLA G, CUOMO S, et al. Unsupervised learning on multimedia data: a cultural heritage case study. Multimedia tools and application, 2020, 79(45/46):34429 – 34442.
[49] REES D, LARAMEE R S. A Survey of Information Visualization Books[J]. Computer graphics forum, 2019, 38(1): 610 – 646.
[50] MANEVITZ L M, FRID A. Cognition and neurocomputation[J]. Annals of mathematics and artificial intelligence, 2020, 88(11 – 12):1119 – 1123.
[51] ZHANG J, ZONG C. Neural machine translation: challenges, progress and future[J]. Science china technological sciences,2020,63(10):2028 – 2050.
[52] LAST M, DANON G. Automatic question generation[J]. Wires data mining knowledge discovery, 2020, 10(6): e1382.
[53] 赵阳洋,王振宇,王佩,等. 任务型对话系统研究综述[J]. 计算机学报,2020,43(10):1862 – 1896.
[54] 黄毅,冯俊兰,胡珉,等. 智能对话系统架构及算法[J]. 北京邮电大学学报,2019,42(6):10 – 19.
[55] LUNTOVSKYY A, GLOBA L, SHUBYN B. From big data to smart data: the most effective approaches for data analytics[M]// ILCHENKO M, URYVSKY L, GLOBA L. Advances in information and communication technology and systems. Cham: Springer, 2020:23 – 40.
[56] FERRAG M A, MAGLARAS L, DERHAB A, et al. Authentication schemes for smart mobile devices: threat models, countermeasures, and open research issues[J]. Telecommunication systems, 2020, 73(2): 317 – 348.
[57] IBTISSAM B,SAMIRA D, BOUABID E O, et al. Using genetic algorithm to improve classification of imbalanced datasets for credit card fraud detection[C]// KHOUKHI F, BAHAJ M, EZZIYYANI M. Proceedings of the International Conference on Advanced Information Technology, Services and Systems (AIT2S-18). Cham: Springer, 2019: 220 – 229.
[58] ISSAOUI Y, KHIAT A, BAHNASSE A, et al. Toward smart logistics: engineering insights and emerging trends [J/OL]. [2020 – 12 – 24]. https://rd. springer. com/article/10. 1007%2Fs11831-020-09494-2.
[59] AHONEN T. Enablers and barriers of smart data-based asset management services in industrial business networks[C]//MATHEW J, LIM C, MA L,et al. Proceedings of the 12th world congress on engineering asset management and the 13th international conference on vibration engineering and technology of machinery. Cham: Springer, 2019: 51 – 60.
[60] NUAIMI E A,NEYADI H A,MOHAMED N,et al. Applications of big data to smart cities[J/OL]. [2020 – 12 – 24]. https://jisajournal. springeropen. com/articles/10. 1186/s13174-015-0041-5.
[61] 刘婷婷,戴慎志,宋海瑜. 智慧社会基础设施新类型拓展与数据基础设施规划编制探索[J]. 城市规划学刊, 2019(4): 95 – 101.
[62] SUKHADIA A, UPADHYAY K, GUNDETI M. et al. Optimization of smart traffic governance system using artificial intelligence[J/OL]. [2020 – 12 – 24]. https://rd. springer. com/article/10. 1007/s41133-020-00035-x.
[63] 马亚雪,李纲,谢辉,等. 数字空间视角下的城市数据画像理论思考[J]. 情报学报, 2019, 38(1): 58 – 67.
[64] ISMAGILOVA E, HUGHES L, RANA N P, et al. Security, privacy and risks within smart cities: literature review and development of a smart city interaction framework[J/OL]. [2020 – 12 – 24]. https://doi. org/10. 1007/s10796-020-10044-1.
[65] BIBRI S E, KROGSTIE J. Environmentally data – driven smart sustainable cities: applied innovative solutions for energy efficiency, pollution reduction, and urban metabolism [J/OL]. [2020 – 12 – 24]. https://energyinformatics. springeropen. com/articles/

10.1186/s42162-020-00130-8.
[66] 糜泽花,钱爱兵.智慧医疗发展现状及趋势研究文献综述[J].中国全科医学, 2019, 22(3): 366 – 370.
[67] 郑建明,孙红蕾.智慧公共文化服务发展战略[J].图书馆论坛, 2020, 40(9): 13 – 19.
[68] LEI S. Construction of intelligent library self-service and management system of college [C] //YANG C T, PEI Y, CHANG J W. Proceeding of innovative computing. Singapore: Springer, 2020: 341 – 348.
[69] ROUHI R, BERTINI F, MONTESI D. Shared images and camera fingerprinting may lead to privacy issues [M]//BOURLAI T, KARAMPELAS P, PATEL V M. Securing social identity in mobile platforms. Cham: Springer, 2020:3 – 19.
[70] KORZUN D, VARFOLOMEYEV A, YALOVITSYNA S, et al. Semantic infrastructure of a smart museum: toward making cultural heritage knowledge usable and creatable by visitors and professionals[J]. Personal and ubiquitous computing, 2017, 21(2): 345 – 354.
[71] 王云弟,谢笑,谢阳群.数据环境下的数描客体概念探析[J].情报理论与实践, 2020, 43(7): 69 – 74,94.
[72] BHATIA M, KAUR A. Quantum computing inspired framework of student performance assessment in smart classroom[J/OL]. [2021 – 02 – 05] https://doi.org/10.1002/ett.4094.
[73] 宋宁远.面向智慧数据的科学知识图谱构建——以 SciGraph 为例[J].科技与出版, 2017(11): 17 – 19.
[74] ALARIFI A, ALWADAIN A. An optimized cognitive-assisted machine translation approach for natural language processing [J]. Computing, 2020, 102(3): 605 – 622.
[75] MANALIGOD H J T, DIÑO M J S, JO S, et al. Knowledge discovery computing for management[J]. Information technology and management, 2020, 21(2): 61 – 62.
[76] 林丽珊.开发知识产品智慧服务科研一线[J].图书情报工作, 2017, 61(S1): 73 – 76.

作者贡献说明:
张云中:确定选题,提出研究思路和框架,设计研究方案,论文修改;
刘嘉琳:文献对比与分析,论文起草及修改。

Smart Data Research: Concept Discrimination, Value Orientation, Key Technologies and Application Framework

Zhang Yunzhong Liu Jialin
Department of Library, Information and Archives, Shanghai University, Shanghai 200444

Abstract: [**Purpose/significance**] The Smart Data is a new concept in the field of data science under the development of "Smart Earth", which theoretical exploration and practical application are developing rapidly. Combing the cognitive veins of the academic circles, gathering consensus and analyzing differences is of great significance to clarify the theoretical system of smart data and promote the application and development of smart data. [**Method/process**] Based on extensive and in-depth reading of relevant literature in domestic and foreign fields, this study divided Smart Data into four aspects: conceptual connotation, value orientation, key technologies and application framework. Overall, this study summarized three conceptual perspectives, five connotation features, five types of value orientation, three clusters of key technologies and five application areas of smart data through comparison and analysis. [**Result/conclusion**] The study found that the essence of smart data lies in its canonical structure and value-added process. Smart data comes from data evolution or structural design, and this enriches its value orientation, making itself presents diversified composite value characteristics. Its technical system support the step-by-step "computable-understandable-conversational" data evolution , the core of its application framework lies in the precise realization of "data" and "user" intelligent interaction. In the future, the theoretical system of smart data still needs to be improved under the view of big "data science", centering on theoretical system construction, data rights governance, balanced development of technology, upgrading service levels, integrating theory and practice, etc.

Keywords: smart data data science value key technology application review